

CLAIMS

We claim:

1. A method for optimizing a database management system process of a query, the method comprising:

5 collecting a plurality of single column statistics for a plurality of columns, the plurality of single column statistics providing an estimate of row counts and unique entry counts for a single column operator;
selecting a preferred single column statistic from the plurality of single column statistics according to a predetermined criteria;
10 storing the preferred single column statistic;
determining a selectivity estimate for predicates in the query using the preferred single column statistic, the selectivity estimate being used in optimizing processing of the query by the database management system.

2. The method of claim 1, wherein the predetermined criteria is a maximum of unique entry counts.

3. The method of claim 2, further comprising:
determining a cross product from the single column statistics; and
calculating the selectivity estimate as the division of the cross product and the
20 maximum of unique entry counts.

4. The method of claim 1, wherein the plurality of single column statistics are selectivities.

25 5. The method of claim 4, wherein the predetermined criteria is a minimum of selectivities.

6. The method of claim 5, further comprising:
determining a cross product from the single column statistics; and
calculating the selectivity estimate as the product of the minimum of selectivities
30 and the cross product.

7. The method of claim 1, wherein the plurality of columns are dependent on each other.
8. A method for optimizing a database management system process of a query, the method comprising:
 - 5 collecting a plurality of single column statistics for a plurality of columns, the plurality of single column statistics providing an estimate of row counts and unique entry counts for a single column operator;
 - selecting a first preferred single column statistic from the plurality of single column statistics according to a first predetermined criteria;
 - 10 determining a second preferred single column statistic from a first relationship of the single column statistics;
 - storing the first and second preferred single column statistics;
 - determining a selectivity estimate for predicates in the query using the first and second preferred single column statistics, the selectivity estimate being used in optimizing processing of the query by the database management system.
9. The method of claim 8, wherein the first predetermined criteria is a maximum of unique entry counts.
10. The method of claim 8, further comprising:
 - determining a cross product from the single column statistics; and
 - calculating the selectivity estimate as the division of the cross product and the maximum of unique entry counts.
11. The method of claim 8, wherein the first relationship of the single column statistics is a product of single column statistics.
12. The method of claim 8, wherein the plurality of single column statistics are selectivities.
13. The method of claim 12, further comprising:
 - determining a cross product from the single column statistics; and

calculating the selectivity estimate as the product of the minimum of selectivities and the cross product.

14. The method of claim 12, wherein the first predetermined criteria is a minimum of selectivities.

15. The method of claim 8, wherein the plurality of columns are dependent on each other.

16. The method of claim 8, wherein the selectivity estimate is within a range between the first and second preferred single column statistics.

17. The method of claim 8, wherein the plurality of columns are substantially independent of each other.

18. The method of claim 17, wherein the selectivity estimate is substantially equal to the first preferred single column statistic.

19. The method of claim 8, wherein the columns are substantially dependent on each other.

20. The method of claim 19, wherein the selectivity estimate is substantially equal to the second preferred column statistic.

21. A method for optimizing a database management system process of a query, the method comprising:

collecting a plurality of single column statistics for a plurality of columns, the plurality of single column statistics providing estimates for row counts and unique entry counts for a single column operator;

determining a first selectivity estimate based on an assumption that the columns are substantially independent of each other;

determining a second selectivity estimate based on an assumption that the columns are substantially dependent on each other;

determining a third selectivity estimate for predicates in the query using the first and second selectivity estimates, the selectivity estimate being used in optimizing processing of the query by the database management system.

- 5 22. The method of claim 21, further comprising:
 determining a cross product from the single column statistics;
 determining a measure of dependency; and
 calculating the selectivity estimate as the product of a difference between the first
and second selectivity estimates plus one of the first or the second selectivity estimates.

10

23. The method of claim 21, wherein the plurality of columns are substantially independent on each other.

24. The method of claim 23, wherein the third selectivity estimate is substantially equal to the first selectivity estimate.

25. The method of claim 21, wherein the plurality of columns are dependent on each other.

26. The method of claim 25, wherein the third selectivity estimate is substantially equal to the second selectivity estimate.

27. The method of claim 21, wherein the third selectivity estimate is within a range between the first and second selectivity estimates.

25 28. The method of claim 27, further comprising determining an estimate of a dependency of the columns.

29. The method of claim 28, wherein the estimate of the dependency of the columns is used to determine the third selectivity estimate.

30

30. The method of claim 21, wherein the third selectivity estimate is chosen to be in a central range between the first and second selectivity estimates.

31. A method for optimizing a database management system process of a query, the method comprising:

collecting a plurality of single column statistics for a plurality of columns, the plurality of single column statistics providing estimates for row counts and unique entry counts for a single column operator;

determining a first selectivity estimate based on an assumption that the columns are substantially independent of each other;

determining a first factor as a measure of a skew of the plurality of columns and as a measure of a dependence of the plurality of the columns;

determining a second selectivity estimate for predicates in the query using the first selectivity estimate and the first factor, the second selectivity estimate being used in optimizing processing of the query by the database management system.

32. The method of claim 31, wherein

a product of unique entry count selectivities is calculated from maximum unique entry counts for the plurality of columns,

a product of maximum initial unique entry counts is calculated from maximum initial unique entry counts for the plurality of columns,

a maximum multicolumn unique entry count is selected from multicolumn entry counts for the plurality of columns, and

the first factor is the product of unique entry count selectivities divided by the product of maximum initial unique entry counts divided by the maximum multicolumn unique entry count.

33. The method of claim 31, wherein the plurality of columns are substantially independent on each other.

34. The method of claim 33, wherein the second selectivity estimate is substantially equal to the first selectivity estimate.

35. The method of claim 31, wherein the plurality of columns are dependent on each other.

5

36. The method of claim 31, wherein the second selectivity estimate is a product of the first factor and the first selectivity estimate.

20060927 09:22:00